

# VU Research Portal

## Integer-Valued Autoregressive Models With Survival Probability Driven By A Stochastic Recurrence Equation

Gorgi, Paolo

### ***published in***

Journal of Time Series Analysis  
2018

### ***DOI (link to publisher)***

[10.1111/jtsa.12272](https://doi.org/10.1111/jtsa.12272)

### ***document version***

Publisher's PDF, also known as Version of record

### ***document license***

Article 25fa Dutch Copyright Act

[Link to publication in VU Research Portal](#)

### ***citation for published version (APA)***

Gorgi, P. (2018). Integer-Valued Autoregressive Models With Survival Probability Driven By A Stochastic Recurrence Equation. *Journal of Time Series Analysis*, 39(2), 150-171. <https://doi.org/10.1111/jtsa.12272>

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

ORIGINAL ARTICLE

# INTEGER-VALUED AUTOREGRESSIVE MODELS WITH SURVIVAL PROBABILITY DRIVEN BY A STOCHASTIC RECURRENCE EQUATION

PAOLO GORGI<sup>a,b,\*</sup>

<sup>a</sup> Department of Econometrics and Operations Research, VU University Amsterdam, Amsterdam, The Netherlands

<sup>b</sup> Tinbergen Institute, Amsterdam, The Netherlands

This paper proposes a new class of integer-valued autoregressive models with a dynamic survival probability. The peculiarity of this class of models lies in the specification of the survival probability through a stochastic recurrence equation. The proposed models can effectively capture changing dependence over time and enhance both the in-sample and out-of-sample performance of integer-valued autoregressive models. This point is illustrated through an empirical application to a real-time series of crime reports. Additionally, this paper discusses the reliability of likelihood-based inference for the class of models. In particular, this study proves the consistency of the maximum likelihood estimator and a plug-in estimator for the conditional probability mass function in a misspecified model setting.

Received 28 November 2016; Accepted 10 October 2017

Keywords: Count time series; INAR models; score-driven models; time-varying parameters.

MOS subject classifications: 62M10; 62M20.

## 1. INTRODUCTION

Over the last few years, a growing interest has emerged in the modeling and forecasting of integer-valued time series. It is shown that dynamic models can be amended such that they can account for the discrete nature of the observed time series. The resulting models clearly lead to better descriptions of the features in integer-valued time series. One of the most popular models for time series of counts is the Al-Osh and Alzaid (1987) and McKenzie (1988)'s integer-valued autoregressive (INAR) model, which is specified based on the thinning operator 'o' in Steutel and Van Harn (1979). For a given natural number  $N \in \mathbb{N}$  and  $\alpha \in (0, 1)$ , the thinning operator is defined to satisfy  $\alpha \circ N = \sum_{i=1}^N x_i$ , where  $\{x_i\}_{i=1}^N$  is a sequence of independent Bernoulli random variables with success probability  $\alpha$ . The thinning operator enables the specification of integer-valued time series models in an autoregressive fashion. In fact, INAR models can be seen as a discrete response version of the linear autoregressive model. The first-order INAR model is described by the following equation:

$$y_t = \alpha \circ y_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}, \quad (1)$$

where  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  is an i.i.d. sequence of integer-valued random variables and  $\mathbb{Z}$  denotes the integers. An appealing feature of the INAR model in equation (1) is its interpretation as a death–birth process. From this interpretation, the coefficient  $\alpha$  is also called the survival probability. As in Al-Osh and Alzaid (1987) and McKenzie (1988)'s original formulation, the error term  $\varepsilon_t$  is typically assumed to be Poisson-distributed. Other distributions were also considered in the literature because the Poisson imposes equidispersion, which can be restrictive in practice (Al-Osh and Aly 1992; Jazi *et al.* 2012). Besides the distribution of the error term, the INAR specification in

\* Correspondence to: Paolo Gorgi, VU University Amsterdam, De Boelelaan 1105, NL-1081HV Amsterdam, The Netherlands. E-mail: p.gorgi@vu.nl

equation (1) was generalized in several directions. Among others, Alzaid and Al-Osh (1990) and Jin-Guan and Yuan (1991) extended the first-order INAR model to a general order  $p$ , Kim and Park (2008) considered a signed thinning operator to handle nonstationary series, and Pedeli and Karlis (2011) introduced a bivariate INAR model.

Real-time series data often exhibit changing dynamic behaviors; thus, employing more flexible specifications for the dynamic component of the model can provide a better description of the underlying behavior of the time series and produce better forecasts. The contribution of this paper is in this direction: we introduce a new class of INAR models with time-varying survival probability. The peculiarity of our approach is that the dynamic of the INAR coefficient is specified through a stochastic recurrence equation (SRE) driven by the score of the predictive likelihood. This method allows us to update the survival probability at each time period using the information provided by past elements of the series. Creal *et al.* (2013) and Harvey (2013) recently proposed the use of the score to update time-varying parameters. Since then, their generalized autoregressive score (GAS) framework has been successfully employed to develop dynamic models in econometrics and time series analyses, see for instance, Salvatierra and Patton (2015), Harvey and Luati (2014), and Creal *et al.* (2011).

Zheng *et al.* (2007) and Zheng and Basawa (2008) also consider time variation in the survival probability of INAR models. Zheng *et al.* (2007) specify the survival probability as a sequence of i.i.d. random variables. This approach leads to a more flexible class of conditional distributions, but due to the i.i.d. assumption, it does not provide a dynamic specification of the INAR coefficient. Zheng and Basawa (2008) allow the INAR coefficient to depend on past observations. Their method introduces a dynamic structure and they update the survival probability using past information, as we do in our approach. However, their specification cannot describe smooth changes in the survival probability.

The proposed INAR model may be interpreted as a misspecified filter to approximate the distribution of an unknown data generating process (DGP). Blasques *et al.* (2015) provide a reasoning behind this interpretation by showing that score-driven filters are optimal in reducing the Kullback–Leibler (KL) divergence with respect to an unknown, true DGP. In this direction, we derive some statistical properties of the maximum likelihood (ML) estimator: we prove its consistency in a misspecified setting and show that the conditional predictive probability mass function (pmf) can be consistently estimated through a plug-in estimator. More specifically, we show that this plug-in estimator converges in probability to a pseudo-true pmf that minimizes the KL divergence with the true pmf of the DGP. Finally, we illustrate the practical usefulness of the proposed model through an application to a real time series of crime reports. The results are promising and show how the dynamic survival probability can enhance both the in-sample and out-of-sample performance of INAR models.

The paper proceeds as follows. Section 2 introduces the class of models. Section 3 discusses the consistency of the ML estimation. Section 4 presents the Monte Carlo results for the finite sample behavior of the ML estimator. Section 5 presents the empirical application. Section 6 concludes.

## 2. INAR MODELS WITH A SCORE-DRIVEN COEFFICIENT

### 2.1. The Class of Models

In this section, we extend the class of INAR models in equation (1) by allowing the survival probability  $\alpha$  to change over time. We specify the dynamics of the time-varying coefficient  $\alpha_t$  based on the score framework of Creal *et al.* (2013) and Harvey (2013). The GAS-INAR model is described by the following equations:

$$y_t = \alpha_t \circ y_{t-1} + \varepsilon_t, \quad (2)$$

$$\text{logit } \alpha_{t+1} = \omega + \beta \text{logit } \alpha_t + \tau s_t, \quad (3)$$

where  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  is an i.i.d. sequence of random variables with pmf  $p_\varepsilon(x, \xi)$  for  $x \in \mathbb{N}$ ,  $\xi \in \Xi \subseteq \mathbb{R}^k$ , the vector  $\theta = (\omega, \beta, \tau, \xi)^\top$  is the  $(k+3)$ -dimensional parameter vector to estimate, and  $s_t = s_t(\alpha_t, \xi)$  denotes the score of the predictive log-likelihood  $\partial \log p(y_t | \alpha_t, y_{t-1}, \xi) / \partial \text{logit } \alpha_t$ . The logit function is  $\text{logit}(x) = \log(x/(1-x))$  for any  $x \in (0, 1)$ . Throughout the paper, we use the convention that the set of natural numbers  $\mathbb{N}$  also includes zero. We can

derive the functional form of the predictive likelihood  $p(y_t|\alpha_t, y_{t-1}, \xi)$  as the convolution between the conditional pmf of  $\alpha_t \circ y_{t-1}$  and the pmf of the error term  $\varepsilon_t$ , that is

$$p(y_t|\alpha_t, y_{t-1}, \xi) = \sum_{k=0}^{m_t} p_b(k, y_{t-1}, \alpha_t) p_e(y_t - k, \xi),$$

where  $m_t = \min(y_t, y_{t-1})$  and  $p_b(x, y_{t-1}, \alpha_t)$ ,  $x \in \{0, \dots, y_{t-1}\}$ , is the pmf of a binomial random variable with size  $y_{t-1}$  and success probability  $\alpha_t$ . Finally, we obtain that the score of the predictive log-likelihood  $s_t = s_t(\alpha_t, \xi)$  is given by

$$s_t(\alpha_t, \xi) = \left( \sum_{k=0}^{m_t} p_{kt}(\alpha_t, \xi) \right)^{-1} \left( \sum_{k=0}^{m_t} p_{kt}(\alpha_t, \xi) (k - y_{t-1} \alpha_t) \right), \quad (4)$$

where

$$p_{kt}(\alpha_t, \xi) = \binom{y_{t-1}}{k} \alpha_t^k (1 - \alpha_t)^{y_{t-1}-k} p_e(y_t - k, \xi).$$

The functional form of the score innovation  $s_t$  depends on the specification of the pmf of the error term  $\varepsilon_t$ . In practice, the pmf  $p_e(x, \xi)$  can be chosen to account for the main features observed in the data. For instance, as we consider in the application in Section 5, a negative binomial distribution may be employed instead of a Poisson distribution when the data suggest overdispersion. Alternatively, a zero-inflated Poisson or zero-inflated negative binomial distribution may be employed when dealing with time series with a large number of zeros. We also note that the GAS-INAR model in equations (2) and (3) retains the interpretation of INAR models as death–birth processes. The observed number of elements  $y_t$  alive at time  $t$  is the sum of the number of surviving elements from time  $t - 1$  and the new birth elements  $\varepsilon_t$ . In our dynamic specification, each of the elements alive at time  $t - 1$  has a probability  $\alpha_t$  of surviving at time  $t$ . The proposed class of models is observation-driven since the dynamic probability  $\alpha_t$  is driven solely by past observations. The score  $s_t$  can be seen as the innovation of the dynamic system in equation (3) because  $s_t$  updates  $\alpha_t$  with the new information that becomes available by observing  $y_t$ . The interpretation of  $s_t$  as an innovation is further justified by the fact that its conditional expectation  $E(s_t|y_{t-1}, \alpha_t)$  is equal to zero.

The main advantage of GAS-INAR models compared to standard INAR models is that they can capture different levels of autocorrelation over time. Time variation in the autocorrelation structure is widely explored in the literature on continuous-valued time series but not in that on integer-valued time series. The GAS-INAR model can be a useful means to fill this gap. In the context of GAS models, Delle Monache and Petrella (2017) and Blasques *et al.* (2014b) proposed a class of AR models that feature time-varying coefficients driven by the score of the predictive log-likelihood. The GAS-INAR model can also be seen as an integer-valued version of their model. In practical applications, we can rely on statistical tests to judge whether or not the data contain evidence of time variation in the survival probability. A possible approach is to consider a likelihood ratio test between the standard INAR model and the GAS-INAR model. Alternatively, one can employ Calvori *et al.* (2016)'s parameter instability test to detect possible time variations in the survival probability (see also Harvey and Thiele 2016). This test is basically a Lagrange multiplier test for GAS effects.

## 2.2. Updating Mechanism for the Survival Probability

Although the functional form of the score innovation  $s_t$  is somewhat complicated, how the past information is processed by the score  $s_t$  to update the survival probability is quite intuitive. In the following, we discuss this intuition and provide a better understanding of why the proposed specification can be very effective in capturing

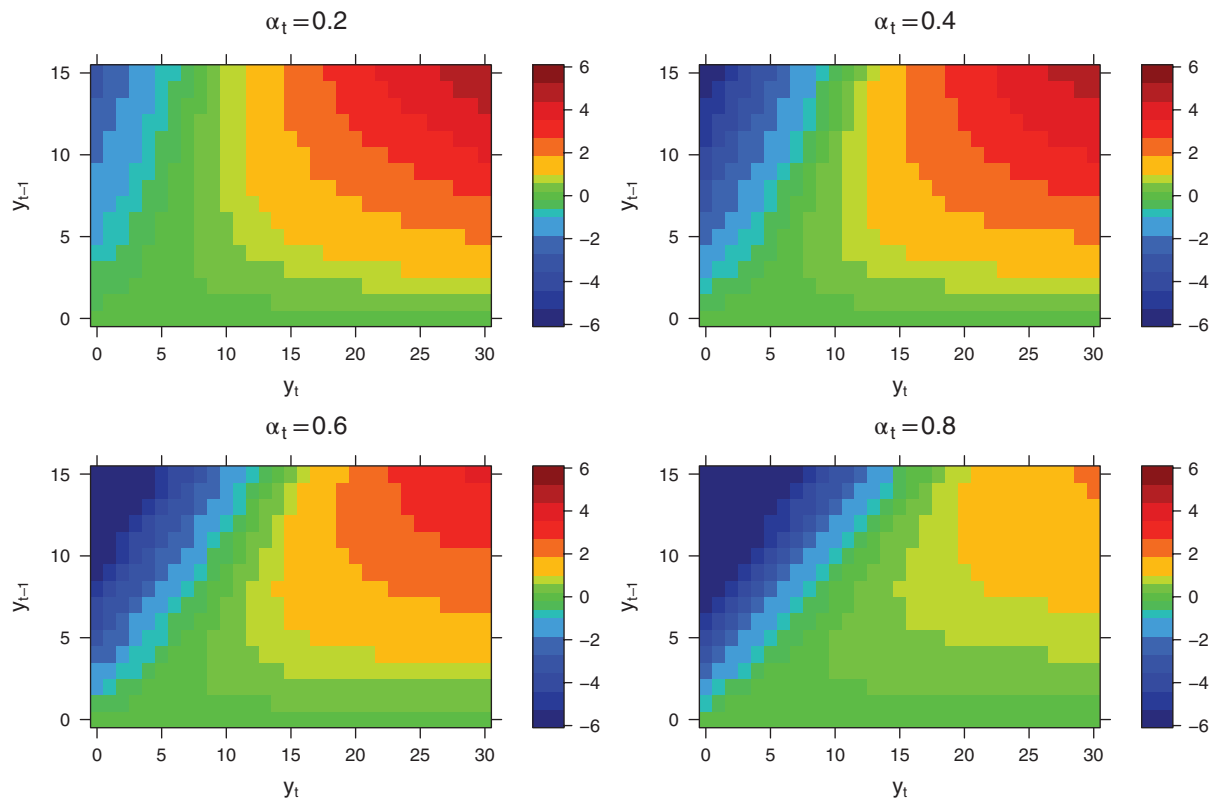


Figure 1. Impact of  $y_t$  and  $y_{t-1}$  on the score innovation  $s_t$  for different values of the survival probability  $\alpha_t$ . A Poisson distribution with mean equal to 5 is considered as distribution of the error term  $\varepsilon_t$ . [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

the changing dependence over time. We illustrate the flexibility of our GAS-INAR model in Section 5 through an empirical application with a real time series of crime data.

It is interesting to see how the information obtained observing  $y_t$  is processed through the score  $s_t$  to update the survival probability from  $\alpha_t$  to  $\alpha_{t+1}$ . Figure 1 describes the impact of  $y_t$  on  $s_t$  for different values of  $y_{t-1}$  and  $\alpha_t$ . As the plots show, the survival probability  $\alpha_t$  gets a negative update when  $y_t$  is small and  $y_{t-1}$  is large. This has an intuitive explanation: the information about  $\alpha_t$  we obtain observing a small  $y_t$  after a large  $y_{t-1}$  is that the survival probability should be small because otherwise, with a large  $\alpha_t$ , we would expect many elements from time  $t - 1$  to survive and thus a large  $y_t$  as well. Consequently,  $\alpha_t$  should get a negative update to discount this information. Similarly, observing a large  $y_t$  following a large  $y_{t-1}$  suggests a high survival probability. Thus, the probability  $\alpha_t$  should be updated accordingly and get a positive innovation  $s_t$ . Finally, an innovation  $s_t$  close to zero may indicate either a lack of information or that the observed value of  $y_t$  is compatible with the value of  $y_{t-1}$  and the current state of the survival probability  $\alpha_t$ . The former case reflects situations in which  $y_{t-1}$  is zero (or close to zero) because observing  $y_t$  gives no information about the survival probability of elements  $y_{t-1}$  since there are no elements alive at  $t - 1$ . On the other hand, the latter case of observing a value  $y_t$  compatible with  $y_{t-1}$  and  $\alpha_t$  can be seen as the green area that separates the red and the blue areas in Figure 1.

This line of reasoning for the direction of update  $s_t$  is subject to the current value of  $\alpha_t$ . For instance, when  $\alpha_t$  is close to zero, perhaps observing a small  $y_t$  after a large  $y_{t-1}$  is exactly what we would expect. Thus, the score update  $s_t$  may be close to zero in this case. The dependence of the score update  $s_t$  on the current survival probability  $\alpha_t$  appears across the different plots in Figure 1.

### 2.3. Parameter Estimation

We can estimate the static parameter vector  $\theta$  of the GAS-INAR model using ML. The log-likelihood function is available in closed form through a prediction error decomposition, namely

$$\hat{L}_T(\theta) = \frac{1}{T} \sum_{t=1}^T \log p(y_t | \hat{\alpha}_t(\theta), y_{t-1}, \xi).$$

We obtain the filtered survival probability  $\hat{\alpha}_t(\theta)$  recursively using the observed data  $\{y_t\}_{t=1}^T$  with

$$\text{logit } \hat{\alpha}_{t+1}(\theta) = \omega + \beta \text{logit } \hat{\alpha}_t(\theta) + \tau s_t(\hat{\alpha}_t(\theta), \xi), \quad (5)$$

where the recursion is initialized at a fixed point  $\text{logit } \hat{\alpha}_0(\theta) \in \mathbb{R}$ . Here, we assume that  $y_{-1}$  and  $y_0$  are given quantities. As we discuss in Section 3, the choice of the initialization  $\hat{\alpha}_0(\theta)$  is irrelevant asymptotically even when the model is misspecified. However, in small samples, using a good starting value has some benefits. A reasonable choice for the initialization is  $\text{logit } \hat{\alpha}_0(\theta) = \omega/(1 - \beta)$ , that is, the unconditional mean  $E\text{logit } \alpha_t$  implied by the GAS-INAR model under the parametric assumption  $\theta$ . This follows immediately because the expected value of the score is equal to zero under standard regularity conditions. The ML estimator is finally given by

$$\hat{\theta}_T = \arg \sup_{\theta \in \Theta} \hat{L}_T(\theta), \quad (6)$$

where  $\Theta$  is a compact parameter set contained in  $\mathbb{R} \times (-1, 1) \times \mathbb{R} \times \Xi$ .

We examine the asymptotic stability of the filtered parameter  $\text{logit } \hat{\alpha}_t(\theta)$  and the consistency of the ML as well as the predictive pmf in Section 3. Furthermore, in Section 4, we perform a simulation experiment to study the finite sample behavior of the ML estimator and to further confirm its reliability.

### 2.4. Forecasting

One of the advantages of properly modeling a count time series by accounting for the discreteness of the data is that users can obtain coherent forecasts of the entire pmf. As Freeland and McCabe (2004a) show, forecasts  $h$  steps ahead are typically available in closed form for standard INAR models. The conditional pmf  $h$  steps ahead can be obtained by repeated applications of the convolution formula. Similarly, for point forecasts, a closed-form expression is available because the conditional expectation  $h$  steps ahead is given by  $E(y_{T+h} | y_T) = \alpha^h y_T + \mu$ , with  $\mu = E(\epsilon_t)$ .

In the following, we illustrate a possible approach to obtain  $h$  steps ahead forecasts for the GAS-INAR model. First, we note that  $y_{T+h}$  depends on the past only through  $\alpha_{T+1}$  and  $y_T$ . Here, we assume that  $\alpha_{T+1}$  and  $y_T$  are given. In practical applications,  $\alpha_{T+1}$  is estimated from the data and  $y_T$  is observed. We define  $p_{T+h|T}(x)$  as the pmf of  $y_{T+h}$  conditional on  $\alpha_{T+1}$  and  $y_T$ . An expression for  $p_{T+h|T}(x)$  is available in closed form only for  $h = 1$ , given by

$$p_{T+1|T}(x) = \sum_{k=0}^{\min\{x, y_T\}} p_b(k, y_T, \alpha_{T+1}) p_e(x - k).$$

Numerical methods are required to obtain  $p_{T+h|T}(x)$  for  $h \geq 2$ . One possibility is to approximate  $p_{T+h|T}(x)$  considering the following simulation scheme. First, we simulate  $B$  realizations for  $y_{T+h}$  conditional on  $\alpha_{T+1}$  and  $y_T$ . We denote these simulations as  $y_{T+h}^i$ , for  $i = 1, \dots, B$ . The recursive scheme described below can be used to obtain the simulations. For each  $i = 1, \dots, B$ , the following two steps are recursively iterated for  $k = 1, \dots, h$ :

- (i) We simulate  $\epsilon_k^i$  from the distribution  $p_e(x, \xi)$  and  $\alpha_{T+k}^i \circ y_{T+k-1}^i$  from a binomial distribution with size  $y_{T+k-1}^i$  and success probability  $\alpha_{T+k}^i$ .

- (ii) We compute  $y_{T+k}^i = \alpha_{T+k}^i \circ y_{T+k-1}^i + \varepsilon_k^i$  and update  $\alpha_{T+k}^i$  to  $\alpha_{T+k+1}^i$  using the following recursion:  $\text{logit } \alpha_{T+k+1}^i = \omega + \beta \text{logit } \alpha_{T+k}^i + \tau s_{t+k}^i$ .

We note that for  $k = 1$ , we have  $\alpha_{T+1}^i = \alpha_{T+1}$  and  $y_T^i = y_T$ . Once we obtain  $B$  draws for  $y_{T+h}$ , we consider an estimate of the conditional pmf given by  $\hat{p}_{T+h|T}(x) = n_x^h/B$ , where  $n_x^h$  denotes the number of draws  $y_{T+h}^i$ ,  $i = 1, \dots, B$ , equal to  $x$ . The next proposition shows that the algorithm above allows us to approximate the model-implied pmf  $p_{T+h|T}(x)$  arbitrarily well by increasing the sample size  $B$  of the simulations.

**Proposition 2.1.** The pmf approximation  $\hat{p}_{T+h|T}(x)$  converges a.s. to the model-implied conditional pmf  $p_{T+h|T}(x)$ , that is,

$$\hat{p}_{T+h|T}(x) \xrightarrow{a.s.} p_{T+h|T}(x) \quad \text{as } B \rightarrow \infty$$

for any  $x \in \mathbb{N}$  and  $h \in \mathbb{N}$ .

Similarly, we can obtain point forecasts  $h$  steps ahead by approximating the conditional expectation  $E(y_{T+h}|y_T, \alpha_{T+1})$  with the sample average  $B^{-1} \sum_{i=1}^B y_{T+h}^i$ . Alternatively, we can use the sample median of  $y_{T+h}^i$ ,  $i = 1, \dots, B$ , to obtain integer forecasts consistent with the discreteness of the data (see Freeland and McCabe (2004a)).

### 3. STATISTICAL PROPERTIES

In this section, we study the reliability of the ML estimation. In the literature, the asymptotic properties of GAS models are often discussed under the assumption of correct specification, see Harvey and Luati (2014) and Harvey (2013), but often also under misspecification, see Bazzi *et al.* (2017) and Blasques *et al.* (2014b). We focus our asymptotic results on model misspecification. We show that the ML estimator is consistent with respect to a pseudo-true parameter that minimizes the KL divergence with an unknown DGP. Consistency arguments with respect to pseudo-true parameters go back to White (1982). In the following, we assume only that the observed data are generated by a stationary and ergodic count process without imposing a specific DGP. Bazzi *et al.* (2017), Blasques *et al.* (2016d), Blasques *et al.* (2016b), Blasques *et al.* (2014b), and Blasques *et al.* (2014a) considered similar assumptions about the DGP.

#### 3.1. Stability of the Filter

A key ingredient to ensure the reliability of the ML estimator for observation-driven models is the stability of the filtered time-varying parameter. The literature typically refers to the stability of the filter as the invertibility of the model (see Straumann and Mikosch 2006; Wintenberger 2013). We first derive the conditions to ensure that the filtered parameter in equation (5) converges to a unique stationary sequence, irrespective of the initialization  $\hat{\alpha}_0(\theta)$ . This result is particularly important because it implies that the initialization is irrelevant asymptotically and provides the basis to ensure the consistency of the ML estimator.

First, we impose some regularity conditions on the pmf of the error term  $p_e(x, \xi)$ .

**Assumption 3.1.** The function  $\xi \mapsto p_e(x, \xi)$  is continuous in  $\Xi$  for any  $x \in \mathbb{N}$  and  $p_e(x, \xi) > 0$  for any  $(x, \xi) \in \mathbb{N} \times \Xi$ .

Assumption 3.1 requires the pmf  $p_e(x, \xi)$  to have full support in  $\mathbb{N}$  and to be continuous with respect to  $\xi$ . These conditions are satisfied for most parametric pmf, such as the Poisson, zero-inflated Poisson, and negative binomial distributions. However, it is worth mentioning that this assumption rules out distributions with limited support, such as the binomial distribution.

The next result ensures stability of the filtered parameter  $\{\hat{\alpha}_t(\theta)\}_{t \in \mathbb{N}}$  specified in equation (5). In particular, Proposition 3.1 shows the exponential almost sure (e.a.s.) uniform convergence of the sequence of functions  $\{\hat{\alpha}_t\}_{t \in \mathbb{N}}$  to a unique stationary and ergodic sequence of functions  $\{\tilde{\alpha}_t\}_{t \in \mathbb{Z}}$ . Note that the filter  $\hat{\alpha}_t$  is a random function that maps from  $\Theta$  into  $(0, 1)$ , whereas the filter  $\hat{\alpha}_t(\theta)$ , evaluated at a given point  $\theta \in \Theta$ , is a random variable. We consider the convergence of the filter with respect to the uniform norm  $\|\cdot\|_\Theta$ , where  $\|f\|_\Theta = \sup_{\theta \in \Theta} |f(\theta)|$  for any function  $f$  that maps from  $\Theta$  into  $\mathbb{R}$ . Recall that a sequence of non-negative random variables  $\{w_t\}_{t \in \mathbb{N}}$  converges e.a.s. to zero if there exists a constant  $\gamma > 1$  such that  $\gamma^t w_t \xrightarrow{a.s.} 0$  as  $t$  diverges.

**Proposition 3.1.** Assume that  $\{y_t\}_{t \in \mathbb{Z}}$  is a stationary and ergodic sequence of count random variables such that  $E y_t^2 < \infty$ . Moreover, let Assumption 3.1 be satisfied, and let the following condition hold:

$$E \log \sup_{\alpha \in (0,1)} |\beta + \tau \dot{s}_t(\alpha, \xi)| < 0, \quad \forall \theta \in \Theta, \quad (7)$$

where  $\dot{s}_t(\alpha, \xi) = \partial s_t(\alpha, \xi) / \partial \log \alpha$  is given by

$$\dot{s}_t(\alpha, \xi) = \frac{\sum_{j=0}^{m_t} \sum_{k=0}^{m_t} p_{kt}(\alpha, \xi) p_{jt}(\alpha, \xi) (k(k-j) - \alpha(1-\alpha)y_{t-1})}{\sum_{j=0}^{m_t} \sum_{k=0}^{m_t} p_{kt}(\alpha, \xi) p_{jt}(\alpha, \xi)}. \quad (8)$$

Then, the filtered parameter  $\{\hat{\alpha}_t(\theta)\}_{t \in \mathbb{N}}$  defined in equation (5) converges e.a.s. and uniformly in  $\Theta$  to a unique stationary and ergodic sequence  $\{\tilde{\alpha}_t(\theta)\}_{t \in \mathbb{Z}}$ , that is,

$$\|\log \hat{\alpha}_t - \log \tilde{\alpha}_t\|_\Theta \xrightarrow{e.a.s.} 0 \quad \text{as } t \rightarrow \infty,$$

for any initialization  $\hat{\alpha}_0$  of the filter.

We provide the proof in Appendix A. Proposition 3.1 does not require a correct specification of the model. The observed data can be generated by any stationary and ergodic count process.

We can check the contraction condition in equation (7) empirically using the observed data; see also the discussion in Blasques *et al.* (2016a). It is impossible to obtain a closed-form expression for equation (7) because it depends on the DGP and the specification of  $p_e(x, \xi)$ . However, with the next proposition we show that the parameter region  $\Theta$  that satisfies equation (7) is not degenerate.

**Proposition 3.2.** The contraction condition (7) of Proposition 3.1 is implied by the following sufficient condition:

$$E \log \max\{|\beta - \tau y_{t-1}/4|, |\beta + \tau m_t^2|\} < 0, \quad \forall \theta \in \Theta,$$

where  $m_t = \min\{y_{t-1}, y_t\}$ .

Proposition 3.2 guarantees that the parameter region  $\Theta$  is not degenerate, as for small enough values of  $|\beta|$  and  $|\tau|$  the inequality is always satisfied.

### 3.2. Consistency of the ML Estimation

We assume that the observed data is a realized path from an unknown DGP  $\{y_t\}_{t \in \mathbb{Z}}$ . Furthermore, we denote with  $p^o(x|y^{t-1})$ ,  $x \in \mathbb{N}$ , the true pmf of  $y_t$  conditional on the past observations  $y^{t-1} = \{y_{t-1}, y_{t-2}, \dots\}$ . The KL divergence



between the true conditional pmf  $p^o(x|y^{t-1})$  and the postulated pmf  $p(x|\tilde{\alpha}_t(\theta), y_{t-1}, \xi)$  is

$$KL_t(\theta) = \sum_{x=0}^{\infty} \log \left( \frac{p^o(x|y^{t-1})}{p(x|\tilde{\alpha}_t(\theta), y_{t-1}, \xi)} \right) p^o(x|y^{t-1}).$$

Note that the conditional KL divergence  $KL_t(\theta)$  depends on  $t$  because it is the KL divergence between the two conditional pmfs at time  $t$ . Therefore,  $KL_t(\theta)$  is a random variable because the conditional pmfs are functions of past observations  $y^{t-1}$ . We are now ready to formally define the pseudo-true parameter  $\theta^*$ .

**Definition 3.1.** The pseudo-true parameter  $\theta^*$  is the minimizer of the average KL divergence  $KL(\theta) = EKL_t(\theta)$  in the parameter set  $\Theta$ .

We also denote with  $\alpha_t^* = \tilde{\alpha}_t(\theta^*)$  the pseudo-true dynamic survival probability and with  $p_t^*(x) = p(x|\alpha_t^*, y_{t-1}, \xi^*)$ ,  $x \in \mathbb{N}$ , the pseudo-true conditional pmf. Below, we also prove the consistency of the plug-in estimators  $\hat{\alpha}_t(\hat{\theta}_T)$  and  $\hat{p}_t(x, \hat{\theta}_T) = p(x|y_{t-1}, \hat{\alpha}_t(\hat{\theta}_T), \hat{\xi}_T)$  for the time-varying survival probability and conditional pmf respectively. This is of practical interest since the main objective of INAR models is typically not the interpretation of the static parameter estimates but approximating the true pmf for forecasting purposes.

We start by considering the following assumption, which imposes some moment conditions and the contraction condition of Proposition 3.1.

**Assumption 3.2.** The following moment conditions hold true:  $Ey_t^2 < \infty$ ,  $E|\log p^o(y_t|y^{t-1})| < \infty$ , and  $E \sup_{\theta \in \Theta} |\log p_e(y_t, \xi)| < \infty$ . Furthermore, the contraction condition in equation (7) is satisfied.

We require Assumption 3.2 to ensure the uniform a.s. convergence of the likelihood function  $\hat{L}_T(\theta)$  to a well-defined deterministic function  $L(\theta) = El_0(\theta)$ , where  $l_t(\theta) = \log p(y_t|\tilde{\alpha}_t(\theta), y_{t-1}, \xi)$  denotes the  $t$ th contribution to the likelihood function when considering the limit filter  $\tilde{\alpha}_t(\theta)$ . Furthermore, we require the integrability condition of the unknown true pmf  $E|\log p^o(y_t|y^{t-1})| < \infty$  to ensure that the average KL divergence exists, and thus the maximizer of  $L(\theta)$  corresponds to the pseudo-true parameter  $\theta^*$ .

Note also that we need the uniform moment condition  $E \sup_{\theta \in \Theta} |\log p_e(y_t, \xi)| < \infty$  only because we consider a general class of pmf for the error term. For most pmfs, this condition is always satisfied. For instance, it holds true immediately as long as  $Ey_t^2 < \infty$  if  $p_e(x, \xi)$  is a Poisson or a negative binomial pmf.

Finally, we impose the following identifiability condition:

**Assumption 3.3.** The function  $L(\theta) = El_0(\theta)$  has a unique maximizer in the set  $\Theta$ .

Assumption 3.3 ensures the uniqueness of the pseudo-true parameter  $\theta^*$ . In general, if this assumption is not satisfied, we find that the limit points of the ML estimator belong to the set of points that minimize the average KL divergence  $KL(\theta)$ .

We are now ready to deliver the strong consistency of the ML estimator with respect to the pseudo-true parameter  $\theta^*$ .

**Theorem 3.1.** Let the observed data  $\{y_t\}_{t=1}^T$  be generated by a stationary and ergodic count process  $\{y_t\}_{t \in \mathbb{Z}}$ ; let the assumptions 3.1, 3.2 and 3.3 be satisfied. Then, the ML estimator defined in equation (6) is strongly consistent with respect to the pseudo-true parameter  $\theta^*$ , that is,

$$\hat{\theta}_T \xrightarrow{\text{a.s.}} \theta^*, \quad T \rightarrow \infty.$$

As a special case of Theorem 3.1, we could also obtain the strong consistency of the ML estimator when the model is correctly specified.

**Remark 3.1.** If we assume that the observed data  $\{y_t\}_{t=1}^T$  are generated by a stationary and ergodic process  $\{y_t\}_{t \in \mathbb{Z}}$  that satisfies the model's equations (2) and (3) for  $\theta = \theta_0$ ,  $\theta_0 \in \Theta$ , it is easy to show that under Assumptions 3.1–3.3, the ML estimator is strongly consistent.

In the next section, we investigate the finite sample properties of the ML estimator under correct specification through a simulation study.

We now turn our attention to examining the consistency of the plug-in estimators  $\hat{\alpha}_t(\hat{\theta}_T)$  and  $\hat{p}_t(x, \hat{\theta}_T)$ . The consistency of these estimators does not follow trivially from the consistency of  $\hat{\theta}_T$  because they are random functions of  $\hat{\theta}_T$  that change at each point in time without converging. Therefore, it is not possible to trivially apply a continuous mapping theorem and immediately obtain the desired consistency. The results we obtain require that both the time index  $t$  and the sample size  $T$  go to infinity because we need  $T \rightarrow \infty$  for the consistency of the ML estimator, that is,  $\hat{\theta}_T \xrightarrow{\text{a.s.}} \theta^*$ , and  $t \rightarrow \infty$  to eliminate the effect of the initialization of the filter, that is,  $\|\text{logit } \hat{\alpha}_t - \text{logit } \tilde{\alpha}_t\|_{\Theta} \xrightarrow{\text{c.a.s.}} 0$ . We note that in practical applications  $t \rightarrow \infty$  is a natural requirement when we are interested in forecasting. The one-step-ahead predictive survival probability  $\hat{\alpha}_{T+1}(\hat{\theta}_T)$  and pmf  $\hat{p}_{T+1}(x, \hat{\theta}_T)$  are obtained by setting  $t = T + 1$ . Therefore, the time index  $t$  is required to grow with the sample size  $T$ , see also Theorem 3 of Wintenberger (2013) for a similar discussion in the context of conditional heteroscedastic models.

The next result shows that the plug-in estimator  $\hat{\alpha}_t(\hat{\theta}_T)$  is strongly consistent with respect to the pseudo-true survival probability  $\alpha_t^*$ .

**Lemma 3.1.** Let the conditions of Theorem 3.1 hold. Then, the plug-in estimator  $\text{logit } \hat{\alpha}_t(\hat{\theta}_T)$  is strongly consistent, that is,

$$\left| \text{logit } \hat{\alpha}_t(\hat{\theta}_T) - \text{logit } \alpha_t^* \right| \xrightarrow{\text{a.s.}} 0, \quad T \rightarrow \infty, \quad t \rightarrow \infty.$$

To determine the consistency of the plug-in estimator  $\hat{p}_t(x, \hat{\theta}_T)$ , we need the following additional regularity condition on the pmf of the error term:

**Assumption 3.4.** The function  $\xi \mapsto p_e(x, \xi)$  is continuously differentiable in  $\Xi$  for any  $x \in \mathbb{N}$ .

Assumption 3.4 is a standard regularity condition that is satisfied by most popular pmfs, such as the Poisson and the negative binomial. The next result delivers consistency in the conditional pmf estimator.

**Theorem 3.2.** Let the observed data  $\{y_t\}_{t=1}^T$  be generated by a stationary and ergodic count process  $\{y_t\}_{t \in \mathbb{Z}}$ ; let the assumptions 3.1–3.4 be satisfied. Then, the conditional pmf plug-in estimator  $\hat{p}_t(x, \hat{\theta}_T)$  is consistent, that is,

$$|\hat{p}_t(x, \hat{\theta}_T) - p_t^*(x)| \xrightarrow{\text{pr.}} 0, \quad T \rightarrow \infty, \quad t \rightarrow \infty$$

for any  $x \in \mathbb{N}$ .

#### 4. FINITE SAMPLE BEHAVIOR OF THE ML ESTIMATOR

We perform a Monte Carlo simulation experiment to test the reliability of the ML estimator in finite samples. We consider the dynamic INAR model specified in equations (2) and (3) with a Poisson error distribution. We denote the mean of the Poisson error  $\varepsilon_t$  by  $\mu$ . The experiment consists of generating 1000 time series of size  $T$  from the GAS-INAR model and estimating the parameter vector  $\theta = (\omega, \beta, \tau, \mu)^\top$  by ML. We consider different parameter values  $\theta$  and sizes  $T$ . Table I provides the simulation results. In particular, Table I reports the mean, bias, standard deviation (SD), and square root of the mean squared error (MSE) of the ML estimator obtained from the 1000 Monte Carlo replications.

Table I. Summary statistics for the sample ML estimator distribution for different parameter values  $\theta$  and sample sizes  $T$ 

		$\omega$	$\beta$	$\tau$	$\mu$	$\omega$	$\beta$	$\tau$	$\mu$
True value		-0.50	0.90	0.15	6.00	-0.50	0.95	0.15	6.00
$T = 250$	Mean	-0.505	0.825	0.161	5.985	-0.496	0.896	0.159	5.996
	Bias	-0.005	-0.075	0.011	-0.015	0.004	-0.054	0.009	-0.004
	SD	0.326	0.175	0.100	0.588	0.411	0.117	0.097	0.570
	$\sqrt{\text{MSE}}$	0.326	0.190	0.101	0.588	0.411	0.129	0.097	0.570
$T = 500$	Mean	-0.496	0.868	0.153	5.986	-0.503	0.927	0.154	5.997
	Bias	0.004	-0.032	0.003	-0.014	-0.003	-0.023	0.004	-0.003
	SD	0.213	0.093	0.062	0.407	0.246	0.053	0.053	0.393
	$\sqrt{\text{MSE}}$	0.213	0.098	0.062	0.407	0.246	0.058	0.053	0.392
$T = 1000$	Mean	-0.494	0.885	0.151	5.987	-0.499	0.939	0.150	5.992
	Bias	-0.006	-0.015	0.001	-0.013	-0.001	-0.011	0.000	-0.008
	SD	0.152	0.050	0.042	0.295	0.171	0.034	0.035	0.279
	$\sqrt{\text{MSE}}$	0.152	0.052	0.042	0.295	0.171	0.036	0.035	0.279
True value		-0.50	0.90	0.30	6.00	-0.50	0.95	0.30	6.00
$T = 250$	Mean	-0.481	0.862	0.304	5.943	-0.502	0.916	0.302	5.945
	Bias	0.019	-0.038	0.004	-0.057	-0.002	-0.034	0.002	-0.055
	SD	0.361	0.095	0.101	0.512	0.501	0.066	0.097	0.473
	$\sqrt{\text{MSE}}$	0.361	0.103	0.101	0.514	0.500	0.075	0.097	0.476
$T = 500$	Mean	-0.495	0.883	0.297	5.971	-0.492	0.935	0.298	5.971
	Bias	0.005	-0.017	-0.003	-0.029	0.008	-0.015	-0.002	-0.055
	SD	0.221	0.044	0.057	0.338	0.361	0.030	0.052	0.310
	$\sqrt{\text{MSE}}$	0.221	0.048	0.057	0.339	0.361	0.033	0.052	0.311
$T = 1000$	Mean	-0.490	0.891	0.299	5.978	-0.502	0.943	0.298	5.981
	Bias	0.010	-0.019	-0.001	-0.022	-0.002	-0.007	0.002	-0.019
	SD	0.156	0.029	0.040	0.242	0.233	0.019	0.035	0.219
	$\sqrt{\text{MSE}}$	0.156	0.031	0.040	0.243	0.233	0.020	0.035	0.220

SD, standard deviation; MSE, mean squared error; ML, maximum likelihood.

The statistics are obtained from 1000 Monte Carlo replications.

The simulation results in Table I further suggest that it is possible to estimate the parameter vector  $\theta$  consistently using ML. This can be elicited from the fact that the MSE of the estimator is decreasing as the sample size  $T$  is increasing. We also note that the estimator of the parameter  $\beta$  tends to be negatively biased in finite samples. In each case, the parameter  $\beta$  is underestimated, on average. The magnitude of the bias also seems to be relevant, especially for  $T = 250$ , because the square root of the MSE is considerably larger than the SD. Therefore, this indicates that the contribution of the bias to the MSE is not negligible compared to the contribution of the variance. The negative bias of  $\beta$  is not surprising since the values of  $\beta$  in the simulations are close to 1 and similar results for the bias are well known in ML estimations of linear autoregressive models. Regarding the other parameters, the results suggest that we can consider the bias as negligible because the SD is almost equal to the square root of the MSE in all considered scenarios.

## 5. APPLICATION TO CRIME DATA

### 5.1. Model Estimation

We present an empirical illustration of the proposed methodology using the monthly number of offensive conduct reports in the city of Blacktown, Australia, from January 1995 to December 2014. We obtained the time series from the New South Wales (NSW) dataset of police reports provided by the NSW Bureau of Crime Statistics and Research and available at <http://www.bocsar.nsw.gov.au/>. The time series is available online as Supporting Information.

Figure 2 provides a plot of the series and shows two periods with a particularly high level of criminal activities: around 2002 and around 2010. During these periods, we expect a higher estimated survival probability  $\alpha_t$ , because they can be seen as periods of high dependence on the past. As Jin-Guan and Yuan (1991) discuss, INAR(p)

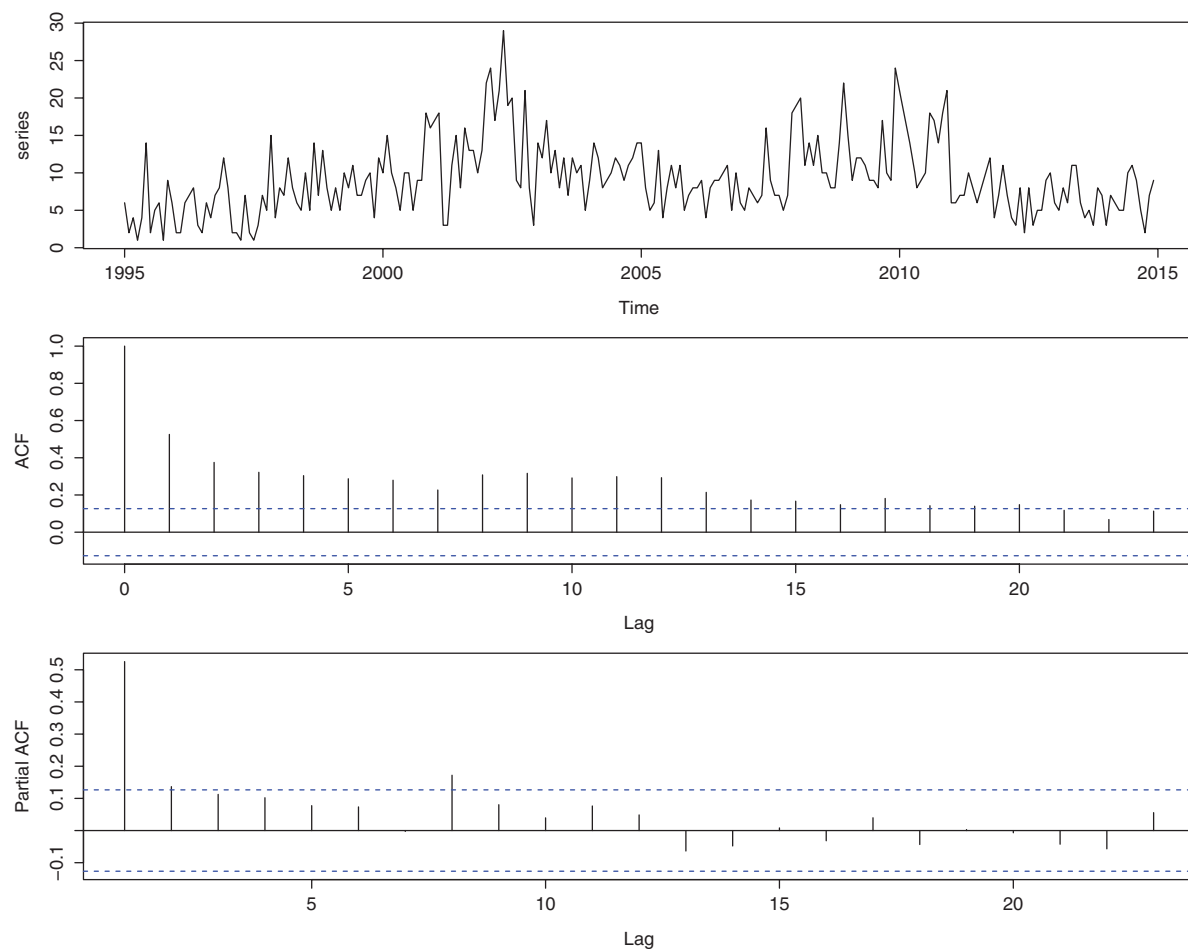


Figure 2. The first plot shows the monthly number of offensive conduct reports in Blacktown from January 1995 to December 2014. The second and third plots represent the sample autocorrelation functions of the series. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

models have the same autocorrelation structure as continuous-valued  $AR(p)$  models. Therefore, we can focus on first-order INAR models because only one lag of the partial autocorrelation function seems to be significant.

We consider several model specifications: the standard INAR models with Poisson and negative binomial error distributions, which we label as PoINAR and NBINAR respectively, and the GAS-INAR models with Poisson and negative binomial error distributions, which we label as GAS-PoINAR and GAS-NBINAR respectively. The sample mean and variance of the data are 9.3 and 24.3 respectively. This indicates overdispersion in the data, and thus a negative binomial distribution for the error term may be more suitable.

Table II presents the ML estimation results. We consider the likelihood ratio test to check the significance of the dynamic coefficient  $\alpha_r$ . Given its meaningful interpretation in a misspecified framework, we also report the Akaike information criterion (AIC) to compare among non-nested models. The results of the likelihood ratio test and the AIC suggest that including the dynamic specification for  $\alpha_r$  plays a relevant role. The likelihood ratio test shows that the dynamic coefficient is highly significant for both the Poisson and the negative binomial specifications. Overall, the model with the smallest AIC is the GAS-NBINAR model. Furthermore, for both negative binomial models, the estimated variance of the error term is more than double the estimated mean. We can thus say that the negative binomial distribution seems to provide a better fit than the Poisson. This result is also consistent with the

Table II. The parameters  $\mu$  and  $\sigma^2$  denote the mean and variance of the error term

	$\omega$	$\beta$	$\tau$	$\mu$	$\sigma^2$	log-lik	$p$ -value	AIC
GAS-NBINAR	-0.907 (0.338)	0.965 (0.027)	0.135 (0.055)	6.083 (0.481)	14.155 (1.853)	-662.91	0.001	1335.82
NBINAR	-0.401 (0.176)	—	—	5.586 (0.456)	15.265 (2.125)	-669.03	—	1344.07
GAS-PoINAR	-1.258 (0.294)	0.967 (0.019)	0.141 (0.033)	6.539 (0.313)	—	-695.04	0.000	1398.24
PoINAR	-0.613 (0.140)	—	—	6.046 (0.323)	—	-714.58	—	1433.21

AIC, Akaike information criterion; GAS, generalized autoregressive score; INAR, integer-valued autoregressive.

The last three columns contain the log-likelihood,  $p$ -value of the likelihood ratio test between the GAS-INAR models and their static INAR counterparts, and the AIC respectively. The asymptotic distribution of the likelihood ratio test is not chi-squared because under the null hypothesis  $\beta$  is not identified. Therefore, the  $p$ -value of the likelihood ratio test is obtained by bootstrap as in Luger (2012). The standard errors are obtained using the inverse of the Fisher information.

overdispersion observed in the data. We can conclude that the results indicate a better in-sample performance for the GAS-INAR model.

From Table II, we also note that the time-varying parameter  $\alpha_t$  is highly persistent because the estimated  $\beta$  is close to 1. Figure 3 plots the estimated path of  $\alpha_t$ , together with the 80% and 95% confidence bounds. As expected, the survival probability is particularly high around 2002 and around 2010. This reflects the high level of criminal activities, which can be interpreted as a higher survival probability of past elements. The plot in Figure 3 also highlights the relevant difference in considering a static  $\alpha$  instead of a dynamic  $\alpha_t$ . We can note this from the fact that the dashed line, which denotes the static parameter estimate of  $\alpha$ , lies outside the 95% confidence bounds of  $\alpha_t$  in some periods. Finally, we also verify the adequacy of the GAS-INAR specification by checking for a residual autocorrelation not captured by the model. We obtain the residuals  $\{\tilde{\epsilon}_t\}_{t=1}^T$  using the formula  $\tilde{\epsilon}_t = y_t - \alpha_t y_{t-1} - \mu$ , as Freeland and McCabe (2004b) propose. Figure 4 reports the autocorrelation functions of the residuals with 95% confidence intervals. We can see that the plot suggests no residual autocorrelation. This further confirms that the GAS-INAR model seems suitable for modeling this time series.

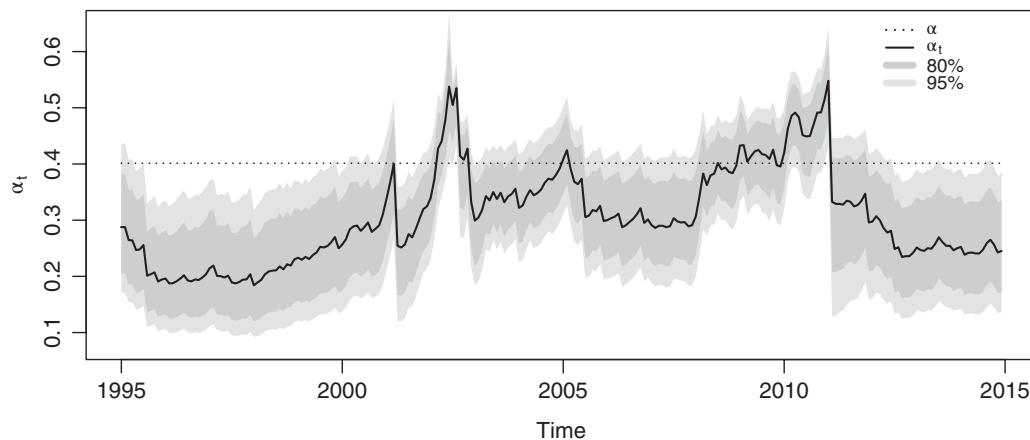


Figure 3. Estimated  $\alpha_t$  from the GAS-NBINAR model with 80% and 95% confidence bounds. The dashed line represents the estimate of  $\alpha$  from the NBINAR model. The confidence bounds are obtained simulating from the distribution of the ML estimator as proposed by Blasques *et al.* (2016c)

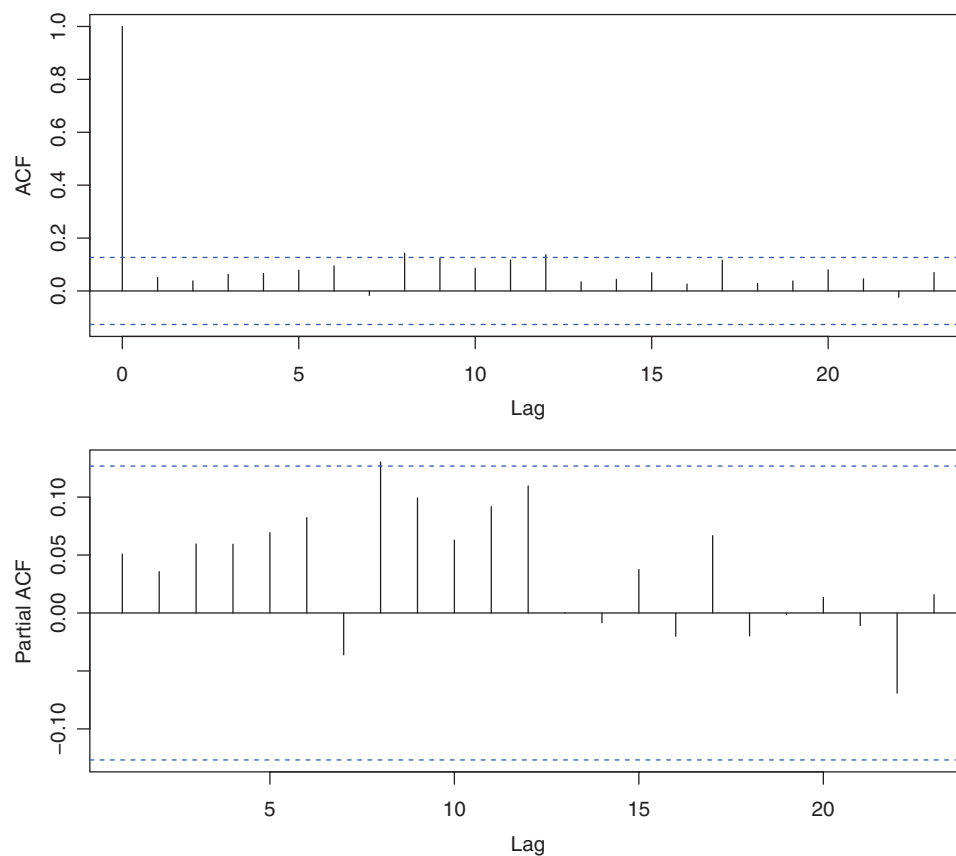


Figure 4. Sample autocorrelation and partial autocorrelation functions of the residuals with 95% confidence bounds. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

## 5.2. Estimation of Other GAS Models

We consider other GAS models for count time series in order to have a benchmark outside the class of INAR models. More specifically, we consider the models of Fokianos *et al.* (2009) and Davis *et al.* (2003), which we label as Po-EINGARCH and Po-INGARCH respectively. Furthermore, we consider a negative binomial specification where the probability  $p$  of the negative binomial is made time-varying through the GAS framework. We label this model NB-INGARCH. The specification of each model is given below.

The Po-INGARCH model is given by

$$y_t | y^{t-1} \sim \mathcal{P}(\mu_t),$$

$$\mu_{t+1} = \omega + \beta \mu_t + \tau(y_t - \mu_t),$$

where  $\mathcal{P}(\mu_t)$  denotes a Poisson distribution with mean  $\mu_t$ . This model is a GAS model with Poisson conditional distribution and identity link function for the time-varying mean  $\mu_t$ . The model is equivalent to the Poisson autoregressive model of Fokianos *et al.* (2009).

Table III. Maximum likelihood estimates of the INGARCH models. Standard errors are in brackets

	$\omega$	$\beta$	$\tau$	$r$	log-lik	AIC
Po-INGARCH	8.755 (0.863)	0.930 (0.041)	0.234 (0.060)	–	–685.58	1377.16
Po-EINGARCH	2.096 (0.152)	0.951 (0.036)	0.206 (0.044)	–	–685.30	1376.60
NB-INGARCH	–0.428 (0.294)	0.954 (0.045)	0.040 (0.013)	12.233 (2.685)	–659.95	1327.90

AIC, Akaike information criterion.

The last two columns contain the log-likelihood and the AIC respectively.

The Po-EINGARCH model is given by

$$y_t | y^{t-1} \sim \mathcal{P}(\mu_t),$$

$$\log \mu_{t+1} = \omega + \beta \log \mu_t + \tau \left( \frac{y_t}{\mu_t} - 1 \right).$$

This model is a GAS model with Poisson conditional distribution and exponential link function for the mean  $\mu_t$ . The model is equivalent to the Poisson model of Davis *et al.* (2003).

Finally, the NB-INGARCH model is given by

$$y_t | y^{t-1} \sim \mathcal{NB}(r, p_t),$$

$$\text{logit } p_{t+1} = \omega + \beta \text{logit } p_t + \tau (y_t(1 - p_t) - rp_t),$$

where  $\mathcal{NB}(r, p_t)$  denotes a negative binomial distribution with success probability  $p_t$  and number of failures equal to  $r$ . The logit link function is considered to ensure that the probability  $p_t$  lies between 0 and 1. This model is a GAS model with negative binomial conditional distribution and logistic link function for the dynamic probability  $p_t$ .

Table III reports the estimate of the models for the time series of crime reports. The results show that also in this case the negative binomial distribution has a better fit than the Poisson distribution. Furthermore, we can see that overall the INGARCH models have a good in-sample performance. This can be noted from the fact that the AIC of the NB-INGARCH model is lower than the AIC of the GAS-NBINAR model; the same is true for the Poisson INGARCH models compared to the GAS-PoINAR.

### 5.3. Out-of-Sample Study

Finally, we perform a pseudo out-of-sample experiment to compare the forecast performances of the models. We split the time series, which is 240 observations, into two subsamples: the first 140 observations, which constitute a training sample, and the last 100 observations as a forecasting evaluation sample. The training sample is then expanded recursively. We evaluate the forecast performance of the models in terms of both point and pmf forecasts. We evaluate the point forecast accuracy by the forecast MSE, that is,  $100^{-1} \sum_{i=1}^{100} (\hat{y}_{T+i} - y_{T+i})^2$ , whereas we evaluate the pmf forecast accuracy by the log score criterion, that is,  $100^{-1} \sum_{i=1}^{100} \log \hat{p}_{T+i|T+i-1}(y_{T+i})$ . The log score criterion provides a means for a comparison based on the KL divergence between the true DGP and the estimated models.

Table IV summarizes the results, which show that including the dynamic survival probability  $\alpha_t$  produces better forecasts. In particular, the GAS-INAR models outperform the standard INAR models in terms of both point and pmf forecasts. Finally, we note that the GAS-INAR models perform well also compared to the INGARCH models. As we can see, the GAS-PoINAR model has a performance similar to that of the Po-INGARCH and Po-EINGARCH models. A similar result occurs for the GAS-NBINAR model compared to the NB-INGARCH model. Overall, we conclude that the GAS-INAR model can be useful in practical applications for forecasting purposes.

Table IV. Forecast MSE and log score criterion computed using the last 100 observations for different forecast horizons  $h$ 

	Mean squared error					
	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$
GAS-NBINAR	<b>15.77</b>	20.17	<b>20.58</b>	<b>21.53</b>	21.34	21.22
NBINAR	16.51	21.43	22.66	23.68	23.88	23.75
GAS-PoINAR	16.33	20.59	21.13	21.99	21.87	21.55
PoINAR	17.00	21.76	22.80	23.82	23.92	23.81
Po-INGARCH	17.25	20.28	20.96	<b>21.53</b>	21.42	20.75
Po-EINGARCH	17.61	<b>20.08</b>	21.00	21.60	<b>21.29</b>	<b>20.47</b>
NB-INGARCH	17.87	21.01	21.54	22.49	22.64	21.53

	Log score criterion					
	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$
GAS-NBINAR	<b>-2.73</b>	<b>-2.81</b>	<b>-2.83</b>	-2.85	<b>-2.85</b>	-2.85
NBINAR	-2.75	-2.85	-2.88	-2.90	-2.90	-2.92
GAS-PoINAR	-2.83	-2.95	-2.99	-2.99	-2.99	-3.00
PoINAR	-2.88	-3.06	-3.11	-3.17	-3.17	-3.18
Po-INGARCH	-2.83	-2.96	-2.95	-2.96	-2.96	-2.92
Po-EINGARCH	-2.84	-2.95	-2.97	-2.99	-2.98	-2.93
NB-INGARCH	-2.75	-2.82	-2.84	<b>-2.84</b>	<b>-2.85</b>	<b>-2.84</b>

## 6. CONCLUSION

In this paper, we have proposed a flexible class of INAR models with dynamic survival probability. These models can be interpreted as misspecified filters to approximate unknown DGPs. They also retain the appealing interpretation of standard INAR models as death–birth processes. The empirical results are promising, as illustrated in the empirical experiment using the time series of crime report data. The GAS-INAR models outperform the standard INAR models both in-sample and out-of-sample. Furthermore, they also have performance similar to those of the INGARCH models in terms of forecasting accuracy, although the in-sample fit seems slightly better for the INGARCH models. Future research may extend the first-order dynamic INAR model to a general order  $p$ . Further work concerns the asymptotic theory of the ML estimator. We have currently only proved the consistency of the estimator. The asymptotic normality requires the study of the first two derivatives of the log-likelihood. In this regard, we have encountered some difficulties in proving the existence of some moments for the derivative processes. This work is planned for future research.

## ACKNOWLEDGEMENTS

The author would like to thank three anonymous referees and a co-editor of the journal for their very constructive comments, which helped in improving the paper.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

## REFERENCES

- Al-Osh MA, Aly E-EA. 1992. First order autoregressive time series with negative binomial and geometric marginals. *Communications in Statistics—Theory and Methods* **21**(9): 2483–2492.
- Al-Osh MA, Alzaid AA. 1987. First-order integer valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis* **8**(3): 261–275.
- Alzaid A, Al-Osh M. 1990. An integer-valued  $p$ th-order autoregressive structure (INAR (p)) process. *Journal of Applied Probability* **27**(2): 314–324.
- Bazzi M, Blasques F, Koopman SJ, Lucas A. 2017. Time-varying transition probabilities for Markov regime switching models. *Journal of Time Series Analysis* **38**(3): 458–478.



- Blasques F, Koopman SJ, Lucas A. 2014a. Maximum likelihood estimation for generalized autoregressive score models. Tinbergen Institute Discussion Paper 14-029/III.
- Blasques F, Koopman SJ, Lucas A. 2014b. Optimal formulations for nonlinear autoregressive processes. Tinbergen Institute Discussion Paper 14-103/III.
- Blasques F, Koopman SJ, Lucas A. 2015. Information-theoretic optimality of observation-driven time series models for continuous responses. *Biometrika* **102**(2): 325–343.
- Blasques F, Gorgi P, Koopman SJ, Wintenberger O. 2016a. Feasible invertibility conditions and maximum likelihood estimation for observation-driven models. Tinbergen Institute Discussion Paper 16-082/III.
- Blasques F, Ji J, Lucas A. 2016b. Semiparametric score driven volatility models. *Computational Statistics & Data Analysis* **100**: 58–69.
- Blasques F, Koopman SJ, Lasak K, Lucas A. 2016c. In-sample confidence bands and out-of-sample forecast bands for time-varying parameters in observation-driven models. *International Journal of Forecasting* **32**(3): 875–887.
- Blasques F, Koopman SJ, Lucas A, Schaumburg J. 2016d. Spillover dynamics for systemic risk measurement using spatial financial time series models. *Journal of Econometrics* **195**(2): 211–223.
- Bougerol P. 1993. Kalman filtering with random coefficients and contractions. *SIAM Journal on Control and Optimization* **31**(4): 942–959.
- Calvori F, Creal D, Koopman SJ, Lucas A. 2016. Testing for parameter instability across different modeling frameworks. *Journal of Financial Econometrics* **15**(2): 223–246.
- Creal D, Koopman SJ, Lucas A. 2011. A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business & Economic Statistics* **29**(4): 552–563.
- Creal D, Koopman SJ, Lucas A. 2013. Generalized autoregressive score models with applications. *Journal of Applied Econometrics* **28**(5): 777–795.
- Davis RA, Dunsmuir WTM, Streett SB. 2003. Observation-driven models for Poisson counts. *Biometrika* **90**(4): 777–790.
- Delle Monache D, Petrella I. 2017. Adaptive models and heavy tails with an application to inflation forecasting. *International Journal of Forecasting* **33**(2): 482–501.
- Fokianos K, Rahbek A, Tjøstheim D. 2009. Poisson autoregression. *Journal of the American Statistical Association* **104**(488): 1430–1439.
- Freeland R, McCabe B. 2004a. Forecasting discrete valued low count time series. *International Journal of Forecasting* **20**(3): 427–434.
- Freeland R, McCabe BP. 2004b. Analysis of low count time series data by Poisson autoregression. *Journal of Time Series Analysis* **25**(5): 701–722.
- Harvey A. 2013. *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. New York: Cambridge University Press.
- Harvey A, Luati A. 2014. Filtering with heavy tails. *Journal of the American Statistical Association* **109**(507): 1112–1122.
- Harvey A, Thiele S. 2016. Testing against changing correlation. *Journal of Empirical Finance* **38**: 575–589.
- Jazi MA, Jones G, Lai C-D. 2012. First-order integer valued AR processes with zero inflated Poisson innovations. *Journal of Time Series Analysis* **33**(6): 954–963.
- Jin-Guan D, Yuan L. 1991. The integer-valued autoregressive (INAR (p)) model. *Journal of Time Series Analysis* **12**(2): 129–142.
- Kim H-Y, Park Y. 2008. A non-stationary integer-valued autoregressive model. *Statistical Papers* **49**(3): 485–502.
- Krengel U. 1985. *Ergodic Theorems*. Berlin: de Gruyter.
- Luger R. 2012. Finite-sample bootstrap inference in GARCH models with heavy-tailed innovations. *Computational Statistics & Data Analysis* **56**(11): 3198–3211.
- McKenzie E. 1988. Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability* **20**: 822–835.
- Pedeli X, Karlis D. 2011. A bivariate INAR (1) process with application. *Statistical Modelling* **11**(4): 325–349.
- Rao RR. 1962. Relations between weak and uniform convergence of measures with applications. *The Annals of Mathematical Statistics* **33**(2): 659–680.
- Salvatierra IDL, Patton AJ. 2015. Dynamic copula models and high frequency data. *Journal of Empirical Finance* **30**: 120–135.
- Steutel F, Van Harn K. 1979. Discrete analogues of self-decomposability and stability. *The Annals of Probability* **7**(5): 893–899.
- Straumann D, Mikosch T. 2006. Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. *The Annals of Statistics* **34**(5): 2449–2495.
- Wald A. 1949. Note on the consistency of the maximum likelihood estimate. *The Annals of Mathematical Statistics* **20**(4): 595–601.
- White H. 1982. Maximum likelihood estimation of misspecified models. *Econometrica* **50**(1): 1–25.
- Wintenberger O. 2013. Continuous invertibility and stable QML estimation of the EGARCH(1,1) model. *Scandinavian Journal of Statistics* **40**(4): 846–867.

Zheng H, Basawa IV. 2008. First-order observation-driven integer-valued autoregressive processes. *Statistics & Probability Letters* **78**(1): 1–9.

Zheng H, Basawa IV, Datta S. 2007. First-order random coefficient integer-valued autoregressive processes. *Journal of Statistical Planning and Inference* **137**(1): 212–229.

## A. APPENDIX

### A.1. Proofs

*Proof of Proposition 2.1.* First, we note that  $y_{T+h}^i, i = 1, \dots, B$ , are independent draws from the distribution of  $y_{T+h}$  given  $\alpha_{T+1}$  and  $y_T$ . This follows from the fact that the proposed algorithm simply simulates paths from the model for some given values of  $\alpha_{T+1}$  and  $y_T$ . Thus, the random variable  $n_x^h$ , which counts the number of draws such that  $y_{T+h}^i = x$ , has a binomial distribution with size  $B$  and success probability  $p_{T+h|T}(x)$ . Therefore, we can conclude that

$$\hat{p}_{T+h|T}(x) = \frac{n_x^h}{B} \xrightarrow{a.s.} p_{T+h|T}(x) \quad \text{as } B \rightarrow \infty$$

by an application of the law of large numbers. □

*Proof of Proposition 3.1.* We derive the convergence result of the proposition on the basis of Theorem 3.1 of Bougerol (1993). Straumann and Mikosch (2006) applied Bougerol's theorem in the space of continuous functions  $\mathbb{C}(\Theta, \mathbb{R})$  equipped with the uniform norm  $\|\cdot\|_\Theta$ . The authors provided stability conditions for functional SRE of the form

$$x_{t+1}(\theta) = \phi_t(x_t(\theta), \theta), \quad t \in \mathbb{N}, \quad (9)$$

where  $x_0(\theta) \in \mathbb{R}$ , the map  $(x, \theta) \mapsto \phi_t(x, \theta)$  from  $\mathbb{R} \times \Theta$  into  $\mathbb{R}$  is almost surely continuous and the sequence  $\{\phi_t(x, \theta)\}_{t \in \mathbb{Z}}$  is stationary and ergodic for any  $(x, \theta) \in \mathbb{R} \times \Theta$ . Wintenberger (2013) weakened Straumann and Mikosch (2006)'s conditions by replacing a uniform contraction condition with a pointwise condition. From Theorem 2 of Wintenberger (2013), we can ensure the uniform e.a.s convergence of a filter that satisfies the SRE in equation (9) if the following conditions hold true:

- (a) There exists an  $x \in \mathbb{R}$  such that  $E \log^+ (\sup_{\theta \in \Theta} |\phi_0(x, \theta)|) < \infty$ ,
- (b)  $E \log^+ (\sup_{\theta \in \Theta} \Lambda_0(\theta)) < \infty$ ,
- (c)  $E \log (\Lambda_0(\theta)) < 0$  for any  $\theta \in \Theta$ ,

where the random coefficient  $\Lambda_t(\theta)$  is given by

$$\Lambda_t(\theta) = \sup_{(x_1, x_2) \in \mathbb{R}^2, x_1 \neq x_2} \frac{|\phi_t(x_1, \theta) - \phi_t(x_2, \theta)|}{|x_1 - x_2|}.$$

For the GAS-INAR model, the random function  $\phi_t$  in equation (9) has the following form:

$$\phi_t(x, \theta) = \omega + \beta x + \tau s_t(\text{logit}^{-1}(x), \xi).$$

We note that  $\phi_t$  satisfies the stationarity and continuity requirements to apply Wintenberger's result. In particular, we obtain that the a.s. continuity of  $\phi_t(x, \theta)$  follows immediately from the a.s. continuity of  $(x, \theta) \mapsto s_t(\text{logit}^{-1}(x), \xi)$ , which is implied by Assumption 3.1, and the continuity of the binomial likelihood (see the functional form of  $s_t$  in equation (4)). Furthermore, the stationarity and ergodicity of  $\{\phi_t\}_{t \in \mathbb{Z}}$  follows from the stationarity and ergodicity of  $\{y_t\}_{t \in \mathbb{Z}}$  together with an application of Proposition 4.3 of Krengel (1985) because

$s_t(\text{logit}^{-1}(x), \xi)$  is a measurable function of  $y_t$  and  $y_{t-1}$ . In the following, we prove the proposition by showing that (a)–(c) are satisfied.

As concerns (a), we set  $x = 0$  and obtain

$$\begin{aligned} E \log^+ \left( \sup_{\theta \in \Theta} |\phi_0(x, \theta)| \right) &\leq \sup_{\theta \in \Theta} |\omega| + \sup_{\theta \in \Theta} |\tau| E \sup_{\theta \in \Theta} |s_t(0.5, \xi)| \\ &\leq \sup_{\theta \in \Theta} |\omega| + \sup_{\theta \in \Theta} |\tau| E |y_{t-1}| < \infty \end{aligned}$$

by an application of Lemma A.1 and because  $E y_0^2 < \infty$ . Thus (a) is satisfied.

As concerns (b), we obtain

$$\begin{aligned} E \log^+ \left( \sup_{\theta \in \Theta} \Lambda_0(\theta) \right) &\leq E \sup_{\theta \in \Theta} \sup_{x \in \mathbb{R}} |\partial \phi_0(x, \theta) / \partial x| \leq \sup_{\theta \in \Theta} |\beta| + \sup_{\theta \in \Theta} |\tau| E \sup_{\theta \in \Theta} \sup_{\alpha \in (0,1)} |\dot{s}(\alpha, \xi)| \\ &\leq \sup_{\theta \in \Theta} |\beta| + \sup_{\theta \in \Theta} |\tau| E |y_{t-1}|^2 < \infty \end{aligned}$$

by an application of Lemma A.1 and because  $E y_0^2 < \infty$ . This shows that (b) holds true.

Finally, as concerns (c), we obtain

$$E \log (\Lambda_0(\theta)) \leq E \sup_{x \in \mathbb{R}} |\partial \phi_0(x, \theta) / \partial x| \leq E \sup_{\alpha \in (0,1)} |\beta + \tau \dot{s}(\alpha, \xi)| < 0$$

for any  $\theta \in \Theta$  by the contraction condition in equation (7). This proves (c) and concludes the proof of the proposition.  $\square$

*Proof of Proposition 3.2.* The result follows immediately by an application of Lemma A.1, which gives an upper bound for the derivative of the score.  $\square$

*Proof of Theorem 3.1.* Assumption 3.3 ensures that  $L(\theta) = E l_t(\theta)$  has a unique maximizer in the compact set  $\Theta$ , which indeed corresponds to the pseudo-true parameter  $\theta^*$  that minimizes  $KL(\theta)$  because  $E |\log p^o(y_t | y^{t-1})| < \infty$  is satisfied by assumption. In the following, we show that the log-likelihood function  $\hat{L}_T(\theta)$  converges almost surely and uniformly in  $\Theta$  to  $L(\theta)$ , that is

$$\|\hat{L}_T - L\|_{\Theta} \xrightarrow{a.s.} 0, \quad T \rightarrow \infty. \quad (10)$$

Then, given the compactness of  $\Theta$  and the identifiability of  $\theta^*$ , the almost sure convergence  $\hat{\theta}_T \xrightarrow{a.s.} \theta^*$  follows by standard arguments due to Wald (1949).

We define  $L_T(\theta) = T^{-1} \sum_{t=1}^T l_t(\theta)$ , with  $l_t(\theta) = \log p(y_t | \tilde{\alpha}_t(\theta), y_{t-1}, \xi)$ , and obtain

$$\|\hat{L}_T - L\|_{\Theta} \leq \|\hat{L}_T - L_T\|_{\Theta} + \|L_T - L\|_{\Theta}, \quad (11)$$

by an application of the triangle inequality. Therefore, the uniform convergence in equation (10) follows if both terms on the right-hand side of the inequality in equation (11) converge almost surely to zero.

First, we show that  $\|\hat{L}_T - L_T\|_{\Theta} \xrightarrow{a.s.} 0$ . An application of the mean value theorem together with Lemma A.1 yields

$$\begin{aligned} |\hat{l}_t(\theta) - l_t(\theta)| &\leq \sup_{\alpha \in (0,1)} |s_t(\alpha, \xi)| |\text{logit } \hat{\alpha}_t(\theta) - \text{logit } \tilde{\alpha}_t(\theta)| \\ &\leq y_{t-1} |\text{logit } \hat{\alpha}_t(\theta) - \text{logit } \tilde{\alpha}_t(\theta)|, \end{aligned}$$

for any  $\theta \in \Theta$  and  $t \in \mathbb{N}$ . Furthermore, taking into account that  $\|\text{logit } \hat{\alpha}_t - \text{logit } \tilde{\alpha}_t\|_{\Theta} \xrightarrow{e.a.s.} 0$  and  $E|y_{t-1}| < \infty$ , an application of Lemma 2.1 of Straumann and Mikosch (2006) yields

$$\sum_{t=1}^{\infty} y_{t-1} \|\text{logit } \hat{\alpha}_t - \text{logit } \tilde{\alpha}_t\|_{\Theta} < \infty,$$

almost surely. As a result, we have that  $T^{-1} \sum_{t=1}^T \|\hat{l}_t - l_t\|_{\Theta} \xrightarrow{a.s.} 0$  and therefore we conclude that the desired result  $\|\hat{L}_T - L_T\|_{\Theta} \xrightarrow{a.s.} 0$  is proved because

$$\|\hat{L}_T - L_T\|_{\Theta} \leq T^{-1} \sum_{t=1}^T \|\hat{l}_t - l_t\|_{\Theta}.$$

We are now left with showing that  $\|L_T - L\|_{\Theta} \xrightarrow{a.s.} 0$ . Note that  $\{l_t\}_{t \in \mathbb{N}}$  is a stationary and ergodic sequence of random elements that take values in the space continuous functions  $\mathbb{C}(\Theta, \mathbb{R})$  equipped with the uniform norm  $\|\cdot\|_{\Theta}$ . Therefore, the desired convergence result follows by an application of the ergodic theorem of Rao (1962) if the uniform integrability condition  $E\|l_t\|_{\Theta} < \infty$  is satisfied. In the following, we show that this condition holds true. First, we note that  $l_t(\theta) \leq 0$  with probability 1 for any  $\theta \in \Theta$  because  $p(y_1|\alpha, y_2, \xi) \leq 1$  for any  $(y_1, y_2, \xi, \alpha) \in \mathbb{N}^2 \times \Xi \times (0, 1)$ . Thus, taking into account that  $\log(1 + \exp(x)) \leq 1 + |x|$  for any  $x \in \mathbb{R}$ , we obtain

$$\begin{aligned} |l_t(\theta)| &= -l_t(\theta) = -\log \sum_{k=0}^{m_t} p_{kt}(\tilde{\alpha}_t(\theta), \xi) \leq -\log p_{0t}(\tilde{\alpha}_t(\theta), \xi) \\ &\leq -y_{t-1} \log(1 - \tilde{\alpha}_t(\theta)) - \log p_e(y_{t-1}, \xi) \\ &\leq y_{t-1} \log(1 + \exp(\text{logit } \tilde{\alpha}_t(\theta))) - \log p_e(y_{t-1}, \xi) \\ &\leq y_{t-1}(1 + |\text{logit } \tilde{\alpha}_t(\theta)|) - \log p_e(y_{t-1}, \xi), \end{aligned}$$

almost surely for any  $\theta \in \Theta$ . Finally, an application of the Cauchy–Schwarz inequality yields

$$E\|l_t\|_{\Theta} \leq Ey_t + Ey_t^2 + E\|\text{logit } \tilde{\alpha}_t\|_{\Theta}^2 + E \sup_{\theta \in \Theta} |\log p_e(y_{t-1}, \xi)| < \infty,$$

where  $Ey_t^2 < \infty$  and  $E \sup_{\theta \in \Theta} |\log p_e(y_{t-1}, \xi)| < \infty$  are satisfied by the assumption and  $E\|\text{logit } \tilde{\alpha}_t\|_{\Theta}^2 < \infty$  follows by an application of Lemma A.2.  $\square$

*Proof of Lemma 3.1.* The proof of this result is an immediate consequence of Theorem 3 of Wintenberger (2013). We simply sketch the main steps to illustrate that all needed conditions are satisfied. The same notation and definitions as in the proof of Proposition 3.1 are considered. First, we note that it is sufficient to show that  $|\text{logit } \tilde{\alpha}_t(\hat{\theta}_T) - \text{logit } \tilde{\alpha}_t^*| \xrightarrow{a.s.} 0$  as  $T \rightarrow \infty$ . This because we have

$$|\text{logit } \hat{\alpha}_t(\hat{\theta}_T) - \text{logit } \tilde{\alpha}_t^*| \leq |\text{logit } \tilde{\alpha}_t(\hat{\theta}_T) - \text{logit } \tilde{\alpha}_t^*| + \|\text{logit } \hat{\alpha}_t - \text{logit } \tilde{\alpha}_t\|_{\Theta},$$

and  $\|\logit \hat{\alpha}_t - \logit \tilde{\alpha}_t\|_{\Theta} \xrightarrow{a.s.} 0$ . Given the results in Theorem 2 of Wintenberger (2013) and the assumptions that are considered in Proposition 3.1, we have that for any  $\theta \in \Theta$  there exists a compact neighborhood  $B(\theta)$  of  $\theta$  such that the contraction condition holds uniformly, that is,  $E \log(\|\Lambda_t\|_{B(\theta)}) < 0$ . Therefore, this is true also for the pseudo-true parameter  $\theta^* \in \Theta$ . As in the proof of Theorem 3 of Wintenberger (2013), repeated applications of the mean value theorem yield

$$\|\logit \tilde{\alpha}_t(\cdot) - \logit \tilde{\alpha}_t^*\|_{B(\theta^*)} \leq \sum_{k=1}^{\infty} \prod_{i=1}^k \|\Lambda_{t-i}\|_{B(\theta^*)} \|\phi_{t-k}(\logit \tilde{\alpha}_{t-k}^*(\cdot) - \logit \tilde{\alpha}_{t-k+1}^*)\|_{B(\theta^*)}$$

with probability 1. The existence of the limit on the right-hand side is obtained by Lemma 2.1 of Straumann and Mikosch (2006) together with the integrability condition  $E \log^+ \|\logit \tilde{\alpha}_t\|_{B(\theta^*)} < \infty$ , which is implied by Lemma A.2, and  $\prod_{i=1}^k \|\Lambda_{t-i}\|_{B(\theta^*)} \xrightarrow{e.a.s.} 0$  as  $k \rightarrow \infty$ . Finally, the desired result  $|\logit \tilde{\alpha}_t(\hat{\theta}_T) - \logit \tilde{\alpha}_t^*| \xrightarrow{a.s.} 0$  follows as in Theorem 3 of Wintenberger (2013) by taking into account that the ML estimator  $\hat{\theta}_T$  is strongly consistent as shown in Theorem 3.1.  $\square$

*Proof of Theorem 3.2.* An application of the mean value theorem together with Lemma A.3 yields that, for any  $x \in \mathbb{N}$ , there is a  $C_x > 0$  and a stationary sequence of random variables  $\{\eta_t\}_{t \in \mathbb{N}}$  such that the following inequalities hold true with probability 1:

$$\begin{aligned} |\hat{p}_t(x, \hat{\theta}_T) - p_t^*(x)| &\leq \sup_{(\alpha, \theta) \in (0,1) \times \Theta} \left| \frac{\partial p(x|y_{t-1}, \alpha, \xi)}{\partial \logit \alpha} \right| |\logit \hat{\alpha}_t(\hat{\theta}_T) - \logit \alpha_t^*| + \\ &\quad + \sup_{(\alpha, \theta) \in (0,1) \times \Theta} \left\| \frac{\partial p(x|y_{t-1}, \alpha, \xi)}{\partial \xi} \right\|_1 \|\hat{\xi}_T - \xi^*\|_1 \\ &\leq \eta_t |\logit \hat{\alpha}_t(\hat{\theta}_T) - \logit \alpha_t^*| + C_x \|\hat{\xi}_T - \xi^*\|_1. \end{aligned}$$

The desired convergence to zero in probability of  $|\hat{p}_t(x, \hat{\theta}_T) - p_t^*(x)|$  then follows immediately as  $\|\hat{\xi}_T - \xi^*\|_1$  is  $o_p(1)$  by Theorem 3.1 and  $|\logit \hat{\alpha}_t(\hat{\theta}_T) - \logit \alpha_t^*|$  is  $o_p(1)$  by Lemma 3.1.  $\square$

## A.2. Technical Lemmas

**Lemma A.1.** Let Assumption 3.1 hold; then the following inequalities are satisfied with probability 1 for any  $\alpha \in (0, 1)$  and  $\xi \in \Xi$ :

- (i)  $|s_t(\alpha, \xi)| \leq 2y_{t-1}$ .
- (ii)  $-y_{t-1}/4 \leq \dot{s}_t(\alpha, \xi) \leq m_t^2$ .

*Proof.* Assumption 3.1 implies that  $p_{kt}(\alpha, \xi) > 0$  with probability 1 for any  $\alpha \in (0, 1)$  and  $\xi \in \Xi$ . This ensures that  $s_t(\alpha, \xi)$  and  $\dot{s}_t(\alpha, \xi)$  are well defined because their denominator, see expressions in equations (4) and (8), is almost surely larger than zero for any  $\alpha \in (0, 1)$  and  $\xi \in \Xi$ .

To show that (i) is satisfied, we note that

$$|s_t(\alpha, \xi)| \leq \left( \sum_{k=0}^{m_t} p_{kt}(\alpha, \xi) \right)^{-1} \left( \sum_{k=0}^{m_t} p_{kt}(\alpha, \xi)(k + y_{t-1}\alpha) \right) \leq (1 + \alpha)y_{t-1}.$$

Therefore, the result in (i) immediately holds true because  $\alpha \in (0, 1)$ .

As concerns (ii), taking into account that  $y_t \geq 0$  almost surely, we obtain that the numerator of the expression in equation (8) is smaller than or equal to

$$\left( \sum_{j=0}^{m_t} \sum_{k=0}^{m_t} p_{kt}(\alpha, \xi) p_{jt}(\alpha, \xi) k(k-j) \right) \leq \left( \sum_{j=0}^{m_t} \sum_{k=0}^{m_t} p_{kt}(\alpha, \xi) p_{jt}(\alpha, \xi) \right) m_t^2.$$

Therefore, it follows immediately that  $\dot{s}_t(\alpha, \xi) \leq m_t^2$ . Similarly, we obtain that the numerator of equation (8) is larger than or equal to

$$\left( \sum_{j=0}^{m_t} \sum_{k=0}^{m_t} p_{kt}(\alpha, \xi) p_{jt}(\alpha, \xi) \right) (-\alpha(1-\alpha)y_{t-1}).$$

Therefore,  $\dot{s}_t(\alpha, \xi) \geq -y_{t-1}/4$  as  $\alpha \in (0, 1)$ . This concludes the proof.  $\square$

**Lemma A.2.** Let the conditions of Proposition 3.1 hold, then  $E\|\text{logit } \tilde{\alpha}_t\|_{\Theta}^2 < \infty$ .

*Proof.* We prove this lemma by showing that there exists a stationary and ergodic sequence  $\{\tilde{v}_t\}_{t \in \mathbb{Z}}$  such that  $E\tilde{v}_t^2 < \infty$  and that  $\|\text{logit } \tilde{\alpha}_t\|_{\Theta} < (\tilde{v}_t + 1)$  with probability 1. Then, it is immediate to conclude that  $E\|\text{logit } \tilde{\alpha}_t\|_{\Theta}^2 < \infty$ . First, we define the sequence  $\{\hat{v}_t\}_{t \in \mathbb{N}}$  through the following SRE:

$$\hat{v}_{t+1} = \omega_u + \beta_u \hat{v}_t + 2\tau_u y_t, \quad t \in \mathbb{N},$$

which is initialized at  $\hat{v}_0 = \omega_u/(1 - \beta_u)$  and where  $\omega_u = \sup_{\theta \in \Theta} |\omega|$ ,  $\beta_u = \sup_{\theta \in \Theta} |\beta|$ , and  $\tau_u = \sup_{\theta \in \Theta} |\tau|$ . Taking into account that  $\beta_u < 1$ , which is implied by the definition of  $\Theta$ , and that  $\{y_t\}_{t \in \mathbb{Z}}$  is stationary and ergodic, an application of Theorem 3.1 of Bougerol (1993) yields that  $|\hat{v}_t - \tilde{v}_t| \xrightarrow{a.s.} 0$  as  $t$  goes to infinity, where  $\{\tilde{v}_t\}_{t \in \mathbb{N}}$  is a stationary and ergodic sequence that admits the following representation:

$$\tilde{v}_t = \omega_u/(1 - \beta_u) + 2\tau_u \sum_{k=1}^{\infty} \beta_u^k y_{t-k}.$$

From this expression, it is straightforward to obtain that  $E\tilde{v}_t^2 < \infty$ , together with  $\beta_u < 1$ , entails  $E\tilde{v}_t^2 < \infty$ .

In the following, we show that  $\|\text{logit } \tilde{\alpha}_t\|_{\Theta} < (\tilde{v}_t + 1)$  with probability 1. Without loss of generality, we can assume that the filter  $\{\text{logit } \hat{\alpha}_t(\theta)\}_{t \in \mathbb{N}}$  is initialized at  $\hat{\alpha}_0(\theta) = \omega/(1 - \beta)$ . Now, taking into account that  $\sup_{\theta \in \Theta} |s_t(\alpha, \xi)| < 2y_{t-1}$  a.s. for any  $\alpha \in (0, 1)$  by Lemma A.1, it follows immediately that  $\|\text{logit } \hat{\alpha}_t\|_{\Theta} \leq \hat{v}_t$  with probability 1 for any  $t \in \mathbb{N}$ . Therefore, we have for a large enough  $t \in \mathbb{N}$  with probability 1

$$\|\text{logit } \tilde{\alpha}_t\|_{\Theta} - \tilde{v}_t - 1 \leq \|\text{logit } \hat{\alpha}_t\|_{\Theta} - \hat{v}_t - 1 + \|\text{logit } \tilde{\alpha}_t - \text{logit } \hat{\alpha}_t\|_{\Theta} + |\tilde{v}_t - \hat{v}_t| < 0$$

because  $\|\text{logit } \tilde{\alpha}_t - \text{logit } \hat{\alpha}_t\|_{\Theta}$  and  $|\tilde{v}_t - \hat{v}_t|$  go to zero almost surely. As a result, given the stationarity of  $\{\|\text{logit } \tilde{\alpha}_t\|_{\Theta} - \tilde{v}_t\}_{t \in \mathbb{Z}}$ , we infer that  $\|\text{logit } \tilde{\alpha}_t\|_{\Theta} < (\tilde{v}_t + 1)$  with probability 1 for any  $t \in \mathbb{Z}$ . This concludes the proof.  $\square$

**Lemma A.3.** Let the conditions of Theorem 3.2 hold. Then, for any  $x \in \mathbb{N}$  there exists a stationary sequence of random variables  $\{\eta_t\}_{t \in \mathbb{N}}$  and a constant  $C_x > 0$  such that almost surely

- (i)  $\sup_{(\alpha, \theta) \in (0, 1) \times \Theta} \left| \frac{\partial p(x|y_{t-1}, \alpha, \xi)}{\partial \text{logit } \alpha} \right| \leq \eta_t.$
- (ii)  $\sup_{(\alpha, \theta) \in (0, 1) \times \Theta} \left\| \frac{\partial p(x|y_{t-1}, \alpha, \xi)}{\partial \xi} \right\|_1 \leq C_x.$

*Proof.* First, we show that (i) holds true. From elementary calculus, we obtain that

$$\frac{\partial p(x|y_{t-1}, \alpha, \xi)}{\partial \log \alpha} = \sum_{k=0}^{m_{xt}} p_{kt}(x, \alpha, \xi)(k - \alpha y_{t-1}),$$

where  $m_{xt} = \min(x, y_{t-1})$  and

$$p_{kt}(x, \alpha, \xi) = \binom{y_{t-1}}{k} \alpha^k (1 - \alpha)^{y_{t-1}-k} p_e(x - k, \xi).$$

As a result, taking into account that  $0 \leq p_{kt}(x, \alpha, \xi) \leq 1$  with probability 1 for any  $(x, \alpha, \xi) \in \mathbb{N} \times (0, 1) \times \Xi$ , it follows that

$$\left| \frac{\partial p(x|y_{t-1}, \alpha, \xi)}{\partial \log \alpha} \right| \leq \sum_{k=0}^{m_{xt}} p_{kt}(x, \alpha, \xi)(k + y_{t-1}) \leq \sum_{k=0}^{y_{t-1}} (k + y_{t-1}) \leq 2(1 + y_{t-1})y_{t-1}.$$

Therefore, the result in (i) is proved by setting  $\eta_t = 2(1 + y_{t-1})y_{t-1}$  and recalling that  $\{y_t\}_{t \in \mathbb{Z}}$  is stationary and ergodic and thus  $\{\eta_t\}_{t \in \mathbb{Z}}$  is stationary and ergodic as well.

As concerns (ii), we have that

$$\frac{\partial p(x|y_{t-1}, \alpha, \xi)}{\partial \xi} = \sum_{k=0}^{m_{xt}} \binom{y_{t-1}}{k} \alpha^k (1 - \alpha)^{y_{t-1}-k} \frac{\partial p_e(x - k, \xi)}{\partial \xi}.$$

As a result, we obtain that the following inequalities are satisfied almost surely:

$$\left\| \frac{\partial p(x|y_{t-1}, \alpha, \xi)}{\partial \log \alpha} \right\|_1 \leq \sum_{k=0}^{m_{xt}} \binom{y_{t-1}}{k} \alpha^k (1 - \alpha)^{y_{t-1}-k} \left\| \frac{\partial p_e(x - k, \xi)}{\partial \xi} \right\|_1 \leq \sum_{k=0}^x \left\| \frac{\partial p_e(x - k, \xi)}{\partial \xi} \right\|_1.$$

Therefore, from the continuity of the derivative provided by Assumption 3.4 and the compactness of  $\Theta$ , we obtain that for any given  $x - k \in \mathbb{N}$  there is a constant  $C_{kx} > 0$  such that

$$\sup_{\theta \in \Theta} \left\| \frac{\partial p_e(x - k, \xi)}{\partial \xi} \right\|_1 \leq C_{kx}.$$

This shows that the result in (ii) holds because  $C_x = \sum_{k=0}^x C_{kx} < \infty$ . □